IIAR JOURNALS

Innovating Knowledge. Advancing Research.

IIAR Journal of Science, Engineering, and Applied Research

Vol 1, No 1, 2026

Physics-Informed and Data-Driven Modelling in Engineering: A Theoretical Framework for Choosing Between Mechanistic, Machine-Learning and Hybrid Approaches Under Data Scarcity

Atolagbe Umar Kawu[1], NWOGBOJI Emmanuel Usulor[2], Paul Idowu Bamigbola[3]

[1,2,3] Veristzon Global Research Lab

ABSTRACT

Engineers predicting the behaviour of a physical system can now proceed in three ways: by deriving a model from first principles, by learning one from data, or by combining the two in a hybrid that embeds physical knowledge within a learned model. In practice, the selection among these approaches is often governed by disciplinary convention or by the availability of software rather than by the characteristics of the problem at hand. This paper develops a framework intended to make that selection a reasoned one. Drawing on recent literature in scientific machine learning, physics-informed neural networks, uncertainty quantification and model validation, and on reviews of hybrid modelling across structural, fluid, thermal, process, energy and materials engineering, it argues that the available approaches differ along three dimensions that bear directly on engineering practice: the quantity of data each requires, the reliability of its predictions beyond the conditions represented in that data, and the degree to which its predictions can be explained and justified. The appropriate choice is shown to depend on three properties of the application: how well the governing physics is understood, how much and what kind of data is available, and how costly a prediction error would be under extrapolation. The paper gives particular attention to validation, which cannot take the same form for a mechanistic model as for a learned one, and to engineering practice under data scarcity, a condition common across much of Africa, where it argues that physical knowledge constitutes the principal safeguard against models that appear accurate but are not.

Keywords: physics-informed machine learning; scientific machine learning; hybrid modelling; surrogate models; data scarcity

Physics-Informed and Data-Driven Modelling in Engineering: A Theoretical Framework for Choosing Between Mechanistic, Machine-Learning and Hybrid Approaches Under Data Scarcity

1. INTRODUCTION

Engineering practice has long depended on predictive models. The structural engineer predicts deflection from beam theory, the process engineer sizes a reactor from kinetic and transport equations, and the power engineer forecasts demand from historical load. For most of the discipline's history these models were mechanistic, derived from conservation laws and constitutive relations and calibrated against a limited number of experiments. Over the past decade, machine learning has introduced a second approach: rather than deriving a model, the engineer trains one, allowing a flexible function approximator to learn the relationship between inputs and outputs directly from data. More recently, a third approach has developed between the two, in which physical knowledge, governing equations, constraints, symmetries, or mechanistic sub-models, is embedded within a learned model so that theory and data constrain one another (Meng et al., 2025; Toscano et al., 2025).

These approaches are now represented across almost every branch of engineering. The review literature documents physics-informed and hybrid methods in fluid mechanics (Sharma et al., 2023), structural integrity (Zhu et al., 2023), bridges (Mammeri et al., 2025), civil, transportation and pavement engineering (Daneshvar et al., 2026), manufacturing and design (Kasilingam et al., 2024; Pan et al., 2025), pharmaceutical process engineering (Gaddem et al., 2024), energy systems (Lin et al., 2025), heat transfer (Nazemzadeh et al., 2026), water and wastewater systems (Di Bella et al., 2026), condition monitoring (Wu & Kundu, 2026) and subsurface engineering (Sinha & Dindoruk, 2025). This literature is extensive and expanding rapidly, but it is organised predominantly by method or by application. It records what has been attempted; it offers little guidance on how an engineer should choose among the available approaches for a given problem.

The present paper addresses that choice. Its central question is which type of model an engineer should adopt for a particular problem, given the state of physical knowledge, the data available, and the consequences of prediction error. The question is relevant in all engineering contexts, but it acquires particular importance where data are scarce, since scarcity undermines the central advantage of machine learning while leaving intact the principal limitation of mechanistic modelling, namely the effort required to derive and calibrate it. A substantial proportion of engineering practice in Africa operates under precisely these conditions: sensor networks are sparse, test facilities are limited, and historical records are often incomplete or held in forms that cannot readily be used for modelling. In such settings, an inappropriate modelling choice does more than waste resources; it can generate unwarranted confidence in predictions that fail when they are most needed.

The paper is theoretical rather than empirical. It does not report new experiments or benchmarks, but synthesises recent reviews and methodological studies into a framework for model selection. The argument proceeds from a characterisation of the three modelling traditions and the hybrid space between them, through the properties that distinguish them and the problem features that should determine their use, to the distinct validation requirements of each and, finally, to the implications for engineering practice under data scarcity, where the framework has its greatest practical significance.

2. LITERATURE REVIEW

2.1 The Three Modelling Traditions and the Hybrid Space Between Them

The mechanistic tradition constructs models from first principles. Conservation of mass, momentum and energy, combined with constitutive relations describing the behaviour of particular materials, yields systems of differential equations that are solved analytically or numerically. The principal strength of such models is that they encode understanding: they generalise across conditions because they represent the mechanisms responsible for a system's behaviour, they can be interrogated when a prediction is unexpected, and they can be applied to designs that have not previously been built. Their limitations are equally well established. Derivation requires expertise; certain phenomena, including turbulence, fatigue crack initiation, degradation chemistry and the behaviour of infrastructure under human use, resist complete mechanistic description; high-fidelity numerical solution can be computationally costly; and every such model contains parameters that must be

© IIARJournal. All rights reserved.

calibrated and simplifying assumptions whose consequences are not always known (Lin et al., 2025).

The data-driven tradition proceeds in the opposite manner. A flexible model, a neural network, a gradient-boosted ensemble or a Gaussian process, is trained to reproduce observed input–output relationships without any requirement that it reflect the mechanisms producing them. Its advantage is the capacity to capture complex relationships that resist derivation, at low computational cost once training is complete. Its weaknesses correspond to the strengths of the mechanistic approach: it generally requires large quantities of representative data, it has no inherent reason to respect physical laws, it is difficult to interpret, and its predictions are unreliable outside the conditions represented in its training data (Pan et al., 2025; Zhu et al., 2023).

Between these two traditions lies a range of hybrid approaches, and it is important not to treat hybrid modelling as a single method. In their review of condition monitoring, Wu and Kundu (2026) distinguish families that generalise well beyond that field. Physical knowledge may guide a learned model externally, by shaping its inputs, features or outputs, for example, through features engineered from mechanistic understanding. It may be incorporated into the training objective, as in the physics-informed neural network, which penalises violations of the governing equations alongside mismatch with data (Farea et al., 2024; Luo et al., 2025). It may be built into the model architecture, so that no configuration of the parameters can violate a specified constraint. A further arrangement, common in process engineering, couples a mechanistic model with a data-driven one in series or parallel, the learned component correcting the errors of the mechanistic model or supplying sub-models that it cannot provide (Gaddem et al., 2024; Kasilingam et al., 2024). The position a model occupies within this space determines its data requirement, its extrapolation behaviour and its interpretability, which is why a framework for model selection must treat the hybrid region as a set of distinct options rather than as a single intermediate alternative.

3. CONCEPTUAL FRAMEWORK

3.1 Properties That Distinguish Engineering Models

Three properties separate these approaches in ways that bear on engineering decisions: the data each requires, the reliability of its predictions under extrapolation, and its interpretability. Although the three are related, it is useful to consider them in turn.

The first concerns the quantity and kind of data a model requires to achieve useful accuracy. A mechanistic model draws on data principally for calibration and validation, and the quantity required scales with the number of uncertain parameters rather than with the complexity of the phenomenon. A data-driven model must learn the entire input–output relationship, so its data requirement increases with the complexity of that relationship and the dimension of the input space. The difficulty of small-data problems has become a significant concern in materials science, where the remedies proposed, database construction, high-throughput computation, active learning and transfer learning, are directed either at increasing the effective quantity of data or at reducing a model's dependence on it (Xu et al., 2023). Hybrid approaches address the same difficulty from a different direction, supplying through physics what would otherwise have to be learned. Reduction of the data requirement is among the most frequently cited motivations for physics-informed machine learning (Meng et al., 2025), and reviews in water-system modelling report substantial reductions in the data needed relative to standard neural networks (Di Bella et al., 2026).

The second property concerns a model's behaviour when required to predict outside the conditions represented in its calibration or training data. This is frequently the property on which engineering consequences turn, since engineers routinely design for conditions that have not been observed, including novel geometries, extreme loads, rare events and future climates. Mechanistic models extrapolate as far as their governing equations remain valid, which is often well beyond the range of available data, though not indefinitely. Data-driven models generally extrapolate poorly, because nothing in their structure constrains predictions outside the training domain (Zhu et al., 2023). Hybrid models occupy an intermediate position, and the extent of their advantage depends on how much of the system's behaviour the embedded physics actually governs (Nazemzadeh et al., 2026; Zhu et al., 2023). An important qualification arises here. Li et al. (2025) demonstrate, in materials science, that many evaluations presented as out-of-distribution tests in fact assess interpolation, because the test

© IIARJournal. All rights reserved.

data lie within regions of the representation space already covered by the training data; on tasks that require genuine extrapolation, performance degrades, and increasing the quantity of data or the size of the model yields limited or even adverse effects. The implication for engineering is that a model's demonstrated performance on held-out data provides little assurance about its behaviour under conditions that differ in kind from those it has encountered.

The third property is interpretability, the degree to which a model's predictions can be explained in terms that engineers, regulators and clients can evaluate against established knowledge. Mechanistic models are interpretable by construction, since each term corresponds to a physical process and an unexpected prediction can be traced to its source. Data-driven models are typically opaque, and post hoc explanation methods provide only partial insight. Hybrid models again occupy an intermediate position, their physical components being interpretable and their learned components not (Pan et al., 2025). In engineering, interpretability is significant not only for understanding but for accountability: designs must be justified to those who review them, and failures must be investigated. A model whose predictions cannot be explained is difficult to certify and difficult to learn from when it fails.

3.2 Matching the Model to the Problem

Because the three properties favour different approaches, no single method is preferable in general; the appropriate choice depends on where a problem stands with respect to three features. The first is the security of physical knowledge. This ranges from systems governed by well-established equations with known or measurable parameters, such as elastic structural response, laminar flow and conduction heat transfer, where mechanistic models are strong and learned models add little beyond computational speed, to systems whose behaviour emerges from processes that are poorly understood or too complex to model directly, such as material degradation in aggressive environments, multiphase flow in heterogeneous porous media, and the behaviour of infrastructure under human use, where mechanistic modelling is unavailable and learning is the only option. Between these extremes lies the larger class of systems in which the dominant physics is understood but particular components, a closure relation, a friction law, a degradation rate, are not. It is in this intermediate region that hybrid approaches are most valuable, employing the known physics where it applies and learning the remainder (Gaddem et al., 2024; Kasilingam et al., 2024).

The second feature is the quantity and character of the available data. Quantity is relevant, but so are coverage, whether the data span the conditions of interest, quality, and type, that is, whether the data comprise dense measurements of internal states or sparse observations of outputs. Data-driven methods are strongest where data are abundant and representative of the conditions of interest, and weakest where data are sparse, noisy or confined to a narrow operating range. Hybrid approaches are of particular value where data are sparse but the physics is partly known, since physical constraints can regularise a learning problem that the data alone cannot resolve (Meng et al., 2025; Xu et al., 2023).

The third feature is the cost of a prediction error under extrapolation. Where an occasional large error can be detected and corrected, or has limited consequences, as in short-term operational forecasting or the preliminary screening of candidate designs, computational speed and low data requirements may reasonably dominate the choice. Where an extrapolation failure can cause loss of life, as in the design of a bridge for untested loads, the assessment of structural integrity in safety-critical components, or the evaluation of a dam under a flood exceeding the historical record, extrapolation behaviour and interpretability become decisive, and the case for incorporating physical structure is correspondingly strong.

Considered together, these three features locate a problem within a space of choices. A problem characterised by well-understood physics, sparse data and a high cost of failure indicates a mechanistic model, possibly supported by a learned surrogate for computational speed whose predictions are checked against the mechanistic model. A problem with poorly understood physics, abundant representative data and a low cost of failure indicates a data-driven approach. A problem with partly understood physics, sparse data and a high cost of failure, a combination that describes a considerable share of practical engineering problems, indicates a hybrid approach, with the physical component selected to govern behaviour in the regions where extrapolation is required. The purpose of locating a problem in this way is not to prescribe a single answer but to make the reasons for a choice explicit and to identify choices made by convention that the problem does not support.

© IIARJournal. All rights reserved.

3.3 Validation Cannot Be Uniform Across Approaches

A model's credibility rests on validation, the demonstration that it represents reality adequately for its intended use. The validation requirements of the three traditions differ systematically, and this distinction has not been fully absorbed into machine-learning practice.

Mechanistic models are validated against experiments chosen to test their assumptions across the intended range of use. Because their structure is fixed by physics, a model that agrees with appropriately selected experiments can be trusted, within limits, at conditions between and beyond them. Data-driven models cannot be validated in this manner. Their agreement with held-out data demonstrates only that they interpolate well within the training distribution, and the finding that many apparent out-of-distribution tests in fact assess interpolation (Li et al., 2025) illustrates how readily standard validation can overstate a learned model's reliability. A data-driven model intended to inform an engineering decision therefore requires validation designed to probe the conditions of actual use, including tests constructed specifically to assess extrapolation.

Hybrid models present an additional difficulty. Physics-informed neural networks are known to exhibit characteristic failure modes: the training objective, which balances physical residuals against mismatch with data, can produce loss landscapes that are difficult to optimise; high-order differential operators can corrupt the gradients; and the weighting between objectives is difficult to tune in the absence of validation data or prior knowledge of the solution (Basir, 2023). Remedies have been proposed (Farhani et al., 2025), and the range of architectures and training strategies continues to expand (Toscano et al., 2025). The existence of these failure modes nevertheless carries an important caution: the incorporation of physics into a training objective does not guarantee that the trained model respects it, and whether it does must be verified rather than assumed.

For these reasons, uncertainty quantification is central to validation across all three approaches rather than an optional refinement. Machine-learning methods for uncertainty quantification in engineering systems are now well developed, covering forward and inverse analysis, reliability assessment and resilience (Shi et al., 2025), with dedicated tools for uncertainty in neural differential equations and operators (Zou et al., 2024). A model used for an engineering decision should report not only a prediction but a defensible estimate of its uncertainty, and that estimate should increase as the model is required to predict in regions poorly covered by its data. A model that expresses uniform confidence across all conditions of use should be regarded with corresponding caution.

3.4 Principal Propositions of the Framework

The framework may be summarised in a set of propositions that are open to empirical examination and that serve as principles for practice; the reasoning supporting each has been presented in the preceding sections.

First, the appropriate modelling approach is determined by three properties of the application, the security of the governing physics, the quantity and character of the available data, and the cost of extrapolation failure, and choices made instead on the basis of disciplinary convention or software availability will frequently be misaligned with the problem (Kasilingam et al., 2024; Lin et al., 2025).

Second, hybrid approaches are not uniformly superior but superior under specific conditions: they dominate both mechanistic and data-driven approaches where the physics is partly known, the data are sparse and extrapolation is required, because in this region physical structure supplies what the data cannot and learning supplies what the physics cannot (Gaddem et al., 2024; Meng et al., 2025; Nazemzadeh et al., 2026). The value of physical knowledge within a hybrid model increases with the scarcity of data and with the distance between the training conditions and the conditions of use; where data are abundant and use remains within the training range, physical knowledge contributes little additional accuracy, though it may still contribute interpretability (Xu et al., 2023; Zhu et al., 2023).

Third, standard held-out validation overstates the extrapolation capability of data-driven and hybrid models, because test sets typically assess interpolation; credible validation for engineering use requires tests constructed specifically to represent the conditions of intended use (Li et al., 2025). Moreover, embedding physics in a training objective does not guarantee physical consistency, since physics-informed models have characteristic failure modes and must be checked against the physical constraints they are intended to satisfy

© IIARJournal. All rights reserved.

rather than assumed to satisfy them (Basir, 2023; Farhani et al., 2025).

Fourth, a model used for engineering decisions must quantify its uncertainty, and that uncertainty must increase with distance from the data; models that report uniform confidence across the conditions of use are unsuitable for safety-critical decisions (Shi et al., 2025; Zou et al., 2024).

Fifth, in data-scarce engineering environments the comparative advantage of physical knowledge is greatest. Investment in physics-informed and hybrid modelling, combined with data collection directed deliberately at the conditions that matter, will yield more reliable models than data-driven methods adopted without adaptation and trained on short or unrepresentative local records (Otunla & Ayegboyin, 2026; Xu et al., 2023).

4. ANALYSIS AND DISCUSSION

4.1 Modelling Under Data Scarcity: The African Context

The framework has particular force where data are scarce, a condition under which much African engineering practice operates. Infrastructure is frequently monitored sparsely or not at all; long records of loads, environmental conditions and system performance are uncommon; test facilities for materials and components are limited, and the data they produce are not always archived in forms that support modelling; and meteorological and hydrological networks are thinner than in the regions where most machine-learning methods are developed. Each of these conditions shifts an application toward the data-scarce end of the spectrum and, with it, toward approaches that rely on physical knowledge.

The renewable-energy sector illustrates the point. Otunla and Ayegboyin (2026) develop a hybrid forecasting model for day-ahead hourly solar radiation across eight Nigerian locations, combining statistical and machine-learning approaches using historical meteorological records spanning one to two years. The work is of practical value, and it also exposes the constraint: a record of this length covers only a few seasonal cycles, so any learned component is trained on a narrow sample of the conditions it will subsequently encounter. In such a setting, the framework indicates that the physical component of a hybrid model should carry as much of the predictive burden as possible, here, the astronomy of solar geometry and the physics of atmospheric attenuation, with the learned component confined to the residual that the physics does not capture. Hybrid approaches that combine physical models with climate-model ensembles for the projection of radiation (Nwokolo et al., 2023) demonstrate how physical structure can extend the reach of limited observations.

Infrastructure monitoring presents a comparable situation. Reviews of physics-informed methods for bridges (Mammeri et al., 2025) and for civil, transportation and pavement engineering (Daneshvar et al., 2026) describe approaches that combine physical models of structural and material behaviour with monitoring data, but most of these approaches were developed in settings where monitoring is dense. In African settings, where a bridge may carry no sensors and a road network may possess little systematic condition data, purely data-driven monitoring is frequently not feasible. The framework indicates a preference for physics-informed approaches capable of drawing useful inferences from sparse observation, supported by data collection directed at the conditions most relevant to safety rather than at whatever data are convenient to obtain. The subsurface sector, important to Nigeria's oil and gas industry, provides a further case: measurements there are sparse, costly and indirect, and reviews of physics-informed machine learning for reservoirs and aquifers accordingly embed flow and transport physics within learned models (Sinha & Dindoruk, 2025), placing such problems firmly within the region where hybrid approaches are expected to dominate.

Two further considerations apply in data-scarce settings. The first concerns transfer. A model trained elsewhere on abundant data may be adapted locally through transfer learning (Xu et al., 2023), but its validity in the new setting cannot be assumed: local conditions may fall outside the source training distribution, and the gap between apparent and genuine generalisation (Li et al., 2025) is likely to be widest when a model is transferred across contexts. The second concerns capability. Hybrid modelling requires engineers competent in both the physics and the learning, and such dual competence is scarce everywhere, more so where engineering education has not yet integrated machine learning. The comparative advantage of physical knowledge in data-scarce settings can be realised only by engineers who retain and deepen that knowledge, rather than substituting for it

© IIARJournal. All rights reserved.

imported tools whose limitations they are not positioned to assess.

4.2 Implications

For engineering practice, the framework provides a structured basis for model selection. Before adopting a method, the engineer should establish how well the governing physics is understood, what data are available and whether they cover the conditions of intended use, and what an extrapolation failure would cost. These determinations locate the application within the decision space and indicate which model properties are most significant. Where they indicate a hybrid approach, the engineer should identify which aspects of behaviour the physics governs reliably and design the model so that physics carries that burden (Gaddem et al., 2024; Kasilingam et al., 2024).

For validation and regulation, the framework implies that learned models used in engineering decisions should be held to validation standards matched to their intended use rather than to the conventions of machine-learning research. Such models should be tested on conditions constructed to assess extrapolation (Li et al., 2025), checked for physical consistency rather than assumed to possess it (Basir, 2023), and required to quantify their uncertainty (Shi et al., 2025). Codes and standards bodies will need to develop guidance on the conditions under which learned and hybrid models may be used in design and assessment and on the evidence required to justify their use.

For engineering education, particularly in African universities, the comparative advantage of physical knowledge under data scarcity supports curricula that integrate machine learning with, rather than substitute it for, the physical sciences, strengthening instruction in mechanics, thermodynamics and transport phenomena alongside learning methods and preparing engineers to combine the two (Pan et al., 2025; Toscano et al., 2025). The framework further identifies a research agenda. The magnitude of the gap between apparent and genuine extrapolation for engineering models in practice remains to be established, as do the conditions under which the physical component of a hybrid model improves extrapolation rather than merely constraining interpolation, the appropriate validation of hybrid models when the physics is only partly known, and the forms of targeted data collection that yield the greatest improvement in model reliability per unit of cost. The last of these would be of particular value to African engineering, where resources for data collection are limited and must be used to best effect.

5. CONCLUSION AND RECOMMENDATION

The choice among mechanistic, data-driven and hybrid models is consequential for the reliability of designs, the safety of infrastructure and the efficient use of limited resources. This paper has argued that the choice should be determined by the characteristics of the problem rather than by convention. Each approach can be characterised by its data requirement, its extrapolation behaviour and its interpretability, and each application can be located by how well its physics is understood, how much data is available and how costly an extrapolation failure would be; the appropriate model is the one whose properties correspond to the application's position within this space.

Two conclusions are of particular importance. The first is that validation cannot take a uniform form across approaches: a learned model's performance on held-out data provides little assurance about conditions differing in kind from those it was trained on, and the presence of physics in a training objective does not guarantee physical consistency. Credible use of learned and hybrid models in engineering therefore requires validation designed for the intended use and a defensible quantification of uncertainty. The second is that data scarcity alters the balance of advantage. Where data are scarce, the comparative advantage of physical knowledge is greatest, and hybrid approaches that allow physics to carry the predictive burden while learning addresses specific residual effects are likely to outperform either pure alternative.

For African engineering, these conclusions carry a clear implication. The potential of machine learning will not be realised by adopting data-driven methods developed for data-rich settings and applying them to short and sparse local records. It will be realised by engineers who combine physical understanding with the judicious use of learning, who collect the data most relevant to their problems rather than those most easily obtained, and who validate their models against the conditions under which those models will actually be used.

6. CONTRIBUTION TO EXISTING KNOWLEDGE

The paper contributes a framework for choosing among physics-informed, data-driven and hybrid

© IIARJournal. All rights reserved.

modelling approaches in engineering by matching the model to the problem along the properties that distinguish engineering models, and it argues that validation cannot be uniform across approaches.

It gives engineers and researchers, particularly those working under data scarcity, a principled basis for selecting a modelling approach and an appropriate validation strategy, rather than defaulting to whichever method is currently in favour, and it specifies the conditions under which each approach and its validation are warranted.

7. REFERENCES

Basir, S. (2023). Investigating and mitigating failure modes in physics-informed neural networks (PINNs). Communications in Computational Physics, 33(5), 1240–1269. https://doi.org/10.4208/cicp.oa-2022-0239

Daneshvar, A., Golroo, A., Nejad, F. M., Shiri, A., & Rasti, M. (2026). Physics-informed neural networks in civil, transportation, and pavement engineering: Cross-domain review, benchmarking challenges, and research gaps. Archives of Computational Methods in Engineering, 33(7), 10629–10653. https://doi.org/10.1007/s11831-026-10643-2

Di Bella, A., Raissi, M., Santoro, D., & Roccaro, P. (2026). Physics-informed neural networks in water and wastewater systems: A critical review. Water Research, 293, 125449. https://doi.org/10.1016/j.watres.2026.125449

Farea, A., Yli-Harja, O., & Emmert-Streib, F. (2024). Understanding physics-informed neural networks: Techniques, applications, trends, and challenges. AI, 5(3), 1534–1557. https://doi.org/10.3390/ai5030074

Farhani, G., Dashtbayaz, N. H., Kazachek, A., & Wang, B. (2025). A simple remedy for failure modes in physics informed neural networks. Neural Networks, 183, 106963. https://doi.org/10.1016/j.neunet.2024.106963

Gaddem, M. R., Kim, J., Matsunami, K., Hayashi, Y., Badr, S., & Sugiyama, H. (2024). Roles of mechanistic, data-driven, and hybrid modeling approaches for pharmaceutical process design and operation. Current Opinion in Chemical Engineering, 44, 101019. https://doi.org/10.1016/j.coche.2024.101019

Kasilingam, S., Yang, R., Singh, S. K., Farahani, M. A., Rai, R., & Wuest, T. (2024). Physics-based and data-driven hybrid modeling in manufacturing: A review. Production & Manufacturing Research, 12(1), 2305358. https://doi.org/10.1080/21693277.2024.2305358

Li, K., Rubungo, A. N., Lei, X., Persaud, D., Choudhary, K., DeCost, B., Dieng, A. B., & Hattrick-Simpers, J. (2025). Probing out-of-distribution generalization in machine learning for materials. Communications Materials, 6(1), 9. https://doi.org/10.1038/s43246-024-00731-w

Lin, Y., Tang, J., Guo, J., Wu, S., & Li, Z. (2025). Advancing AI-enabled techniques in energy system modeling: A review of data-driven, mechanism-driven, and hybrid modeling approaches. Energies, 18(4), 845. https://doi.org/10.3390/en18040845

Luo, K., Zhao, J., Wang, Y., Li, J., Wen, J., Liang, J., Soekmadji, H., & Liao, S. (2025). Physics-informed neural networks for PDE problems: A comprehensive review. Artificial Intelligence Review, 58(10), 323. https://doi.org/10.1007/s10462-025-11322-7

Mammeri, S., Barros, B., Conde-Carnero, B., & Riveiro, B. (2025). From traditional damage detection methods to physics-informed machine learning in bridges: A review. Engineering Structures, 330, 119862. https://doi.org/10.1016/j.engstruct.2025.119862

Meng, C., Griesemer, S., Cao, D., Seo, S., & Liu, Y. (2025). When physics meets machine learning: A survey of physics-informed machine learning. Machine Learning for Computational Science and Engineering, 1(1), 20. https://doi.org/10.1007/s44379-025-00016-0

Nazemzadeh, N., Loyola-Fuentes, J., Righetti, G., Diaz-Bejarano, E., Mancin, S., & Coletti, F. (2026). Hybrid modelling of heat transfer systems: Combining physics-based and data-driven approaches for improved prediction and extrapolation. AI Thermal Fluids, 5, 100030. https://doi.org/10.1016/j.aitf.2026.100030

Nwokolo, S. C., Proutsos, N., Meyer, E. L., & Ahia, C. C. (2023). Machine learning and physics-based hybridization models for evaluation of the effects of climate change and urban expansion on photosynthetically active radiation. Atmosphere, 14(4), 687. https://doi.org/10.3390/atmos14040687

Otunla, T. A., & Ayegboyin, E. O. (2026). Day-ahead hourly forecasting of solar radiation using a physics-based hybrid machine learning model in some selected locations in Nigeria. Nigerian Journal of Physics, 35(2), 83–95. https://doi.org/10.62292/njp.v35i2.2026.503

Pan, L., Li, G., Zhu, T., Liu, D., Wang, Y., & Lu, Y. (2025). Physics-informed machine learning in design and manufacturing: Status and challenges. Journal of Computing and Information Science in Engineering, 25(12), 120804. https://doi.org/10.1115/1.4070100

Sharma, P., Chung, W. T., Akoush, B., & Ihme, M. (2023). A review of physics-informed machine learning in fluid mechanics. Energies, 16(5), 2343. https://doi.org/10.3390/en16052343

© IIARJournal. All rights reserved.

Shi, Y., Wei, P., Feng, K., Feng, D.-C., & Beer, M. (2025). A survey on machine learning approaches for uncertainty quantification of engineering systems. Machine Learning for Computational Science and Engineering, 1(1), 11. https://doi.org/10.1007/s44379-024-00011-x

Sinha, U., & Dindoruk, B. (2025). Review of physics-informed machine learning (PIML) methods applications in subsurface engineering. Geoenergy Science and Engineering, 250, 213713. https://doi.org/10.1016/j.geoen.2025.213713

Toscano, J. D., Oommen, V., Varghese, A. J., Zou, Z., Ahmadi Daryakenari, N., Wu, C., & Karniadakis, G. E. (2025). From PINNs to PIKANs: Recent advances in physics-informed machine learning. Machine Learning for Computational Science and Engineering, 1(1), 15. https://doi.org/10.1007/s44379-025-00015-1

Wu, H., & Kundu, P. (2026). Review of physics augmented: Physics-guided, physics-informed, and physics-encoded machine learning models in condition monitoring. Engineering Applications of Artificial Intelligence, 182, 116100. https://doi.org/10.1016/j.engappai.2026.116100

Xu, P., Ji, X., Li, M., & Lu, W. (2023). Small data machine learning in materials science. npj Computational Materials, 9(1), 42. https://doi.org/10.1038/s41524-023-01000-z

Zhu, S.-P., Wang, L., Luo, C., Correia, J. A. F. O., De Jesus, A. M. P., Berto, F., & Wang, Q. (2023). Physics-informed machine learning and its structural integrity applications: State of the art. Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences, 381(2260), 20220406. https://doi.org/10.1098/rsta.2022.0406

Zou, Z., Meng, X., Psaros, A. F., & Karniadakis, G. E. (2024). NeuralUQ: A comprehensive library for uncertainty quantification in neural differential equations and operators. SIAM Review, 66(1), 161–190. https://doi.org/10.1137/22m1518189

© IIARJournal. All rights reserved.